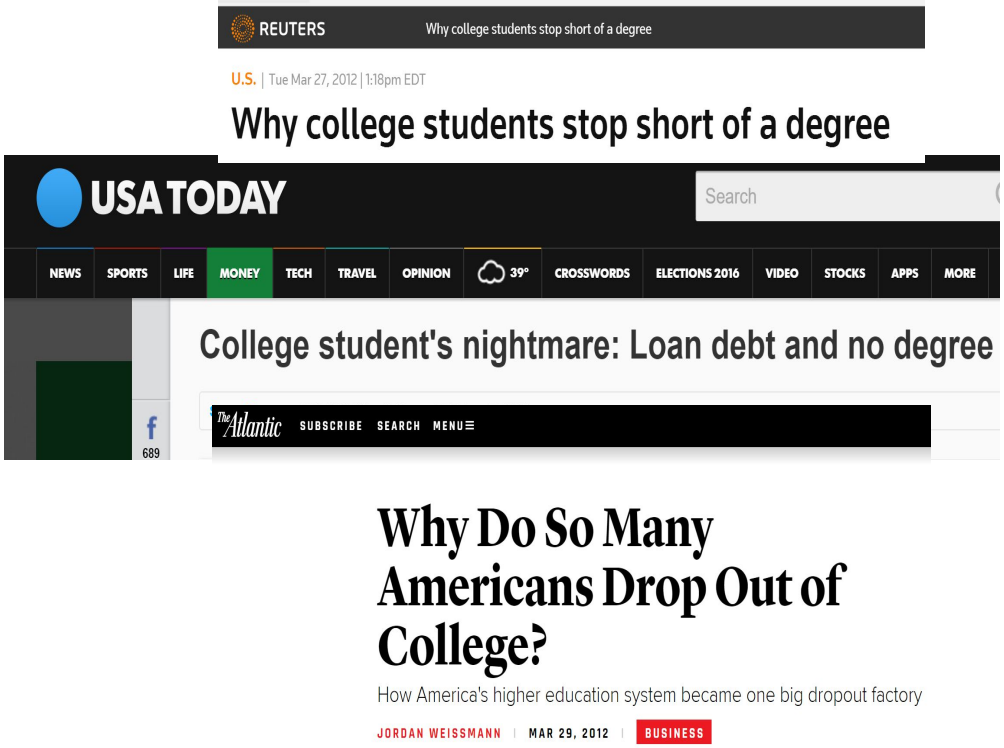
W271 Lab 3: Impact of Financial Aid and Total Cost on Institutional Completion Rates

Shih Yu Chang, Nick Stamatakis, Tony Panza

Overview

About 60% of students who begin studying for a bachelor's degree succeed in attaining a diploma.

Many institutions of higher education are facing pressure from state and federal authorities to increase graduation rates.



Hypothesis: Institutions with lower costs and greater student financial support show higher graduation rates.
This relationship is particularly impactful in schools with large percentages of low incomes students.

Data: College Scorecard and Integrated Postsecondary Education Data System - over 8,000 institutions of higher education, with 7,000 + variables.

EDA: Two exploratory models to gain a greater sense of the relationships:

"Cost Model": $$y_{CompletionRate} = \beta_0 + \beta_1 x_{AttendCost} + \beta_2 x_{AdmissionRate} + \beta_3 x_{SATavgScore} + \beta_4 x_{InstrExpend}$$

"Aid Model": $$y_{CompletionRate} = \beta_0 + \beta_1 x_{FederalGrant} + \beta_2 x_{StateGrant} + \beta_3 x_{Loans}$$

Cost Model

Data Set

8,500 Colleges and Universities

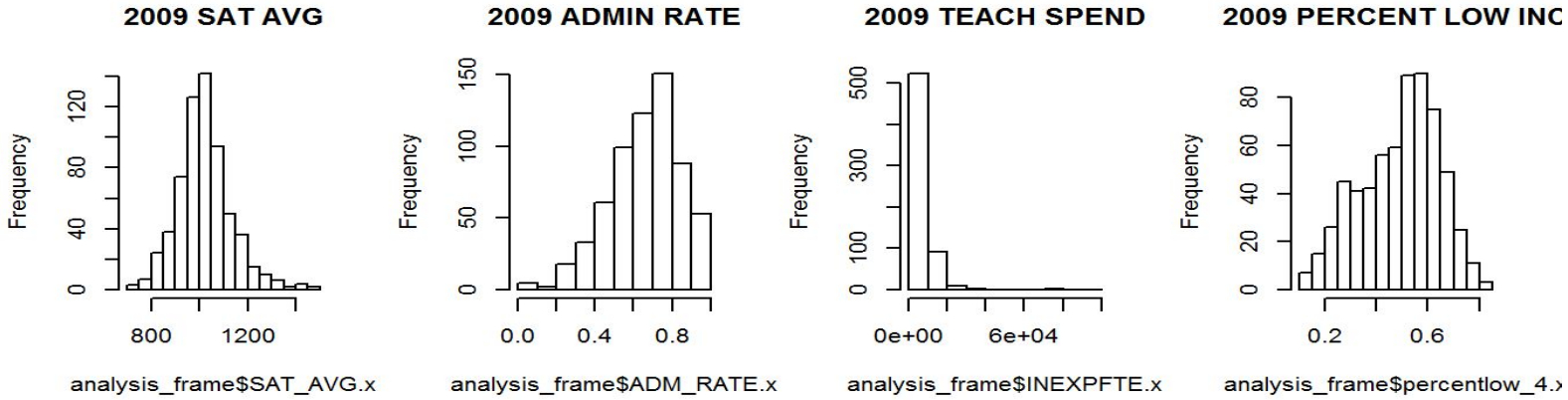
Incomplete Data

Non-bachelors Institutions

No low income breakout

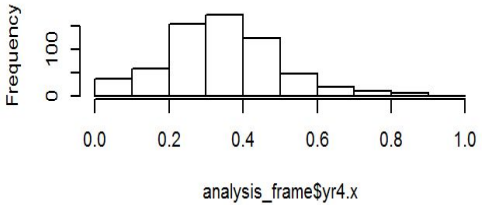
640 Colleges and Universities (similar to US Today list)

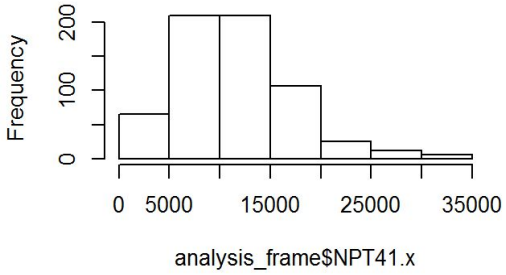
Control Variables

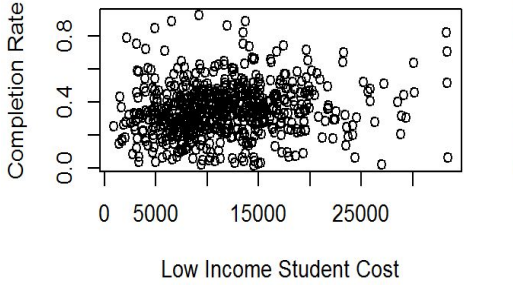


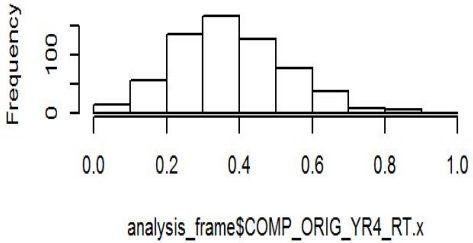
Dependent and Independent Variables

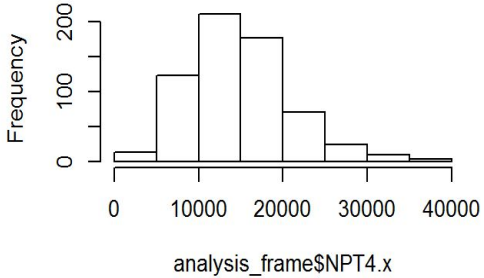
	Y	~	X	X vs. Y
Among Low Income Students	Completion Rate		General Cost	Cost vs. Completion
Among All Students				

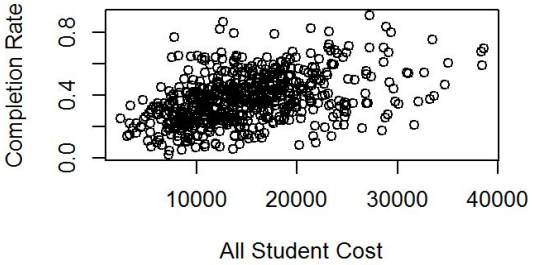








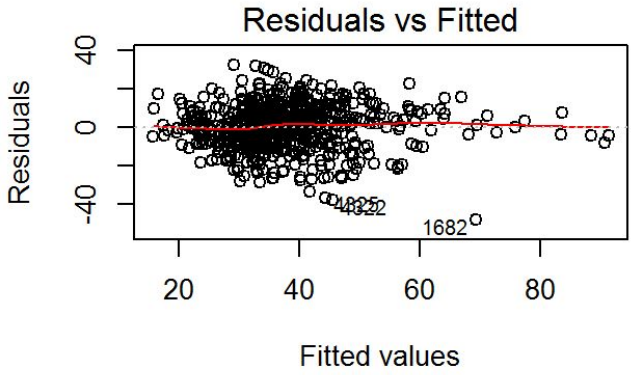


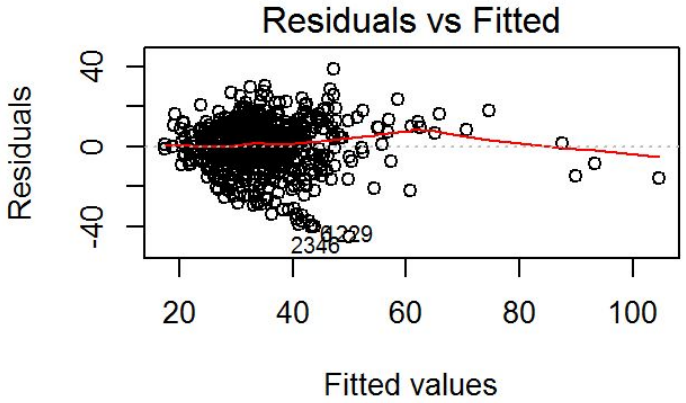


Regression on the Cost Model

```
all_2009 <- lm(COMP_ORIG_YR4_RT.x * 100 ~ NPT4.x + ADM_RATE.x +
SAT_AVG.x + INEXPFTE.x, data = analysis_frame)
coeftest(all_2009, vcov=vcovHC)
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept) -3.4102e+01  4.5645e+00 -7.4712 2.675e-13 ***
## NPT4.x       7.2575e-04  8.4017e-05  8.6381 < 2.2e-16 ***
## ADM_RATE.x  -4.4208e+00  2.5624e+00 -1.7253   0.08497 .
## SAT_AVG.x    6.0116e-02  4.8692e-03 12.3462 < 2.2e-16 ***
## INEXPFTE.x   2.9178e-04  6.2290e-05  4.6843 3.446e-06 ***
```

```
low_2009 <- lm(yr4.x * 100 ~ NPT41.x + ADM_RATE.x + SAT_AVG.x +
INEXPFTE.x + percentlow_4.x, data = analysis_frame)
coeftest(low_2009, vcov=vcovHC)
## t test of coefficients:
##
##                   Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)     1.0895e+00  7.1548e+00  0.1523  0.879017
## NPT41.x         2.5573e-04  1.1301e-04  2.2629  0.023986 *
## ADM_RATE.x     -7.9195e+00  2.7463e+00 -2.8837  0.004065 **
## SAT_AVG.x       3.8699e-02  5.8524e-03  6.6124 8.100e-11 ***
## INEXPFTE.x      5.7649e-04  1.0666e-04  5.4049 9.226e-08 ***
## percentlow_4.x -1.7364e+01  4.0215e+00 -4.3178 1.831e-05 ***
```





Part II: Financial Aid Effect on Undergraduate Completion

Model 2 Hypothesis and Data Source

Data obtained from IPEDS. Used cross-section from 2009 to match Scorecard

However, this did not have graduation rate broken down by income.

Therefore, we decided to formulate and test a different but related hypothesis.

We took a look at financial aid and whether it affects graduation rate. Three categories of financial aid were analyzed:

- Federal grants (`fed_grant_pct`, `fed_grant_avg_amount`)
- State grants (`state_grant_pct`, `state_grant_avg_amount`)
- Loans (`loan_pct`, `loan_avg_amount`)

EDA of IPEDS Data

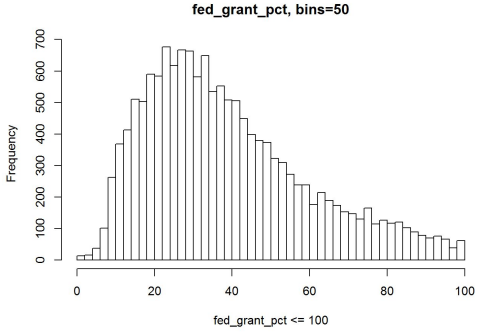
14,761 Colleges and Universities

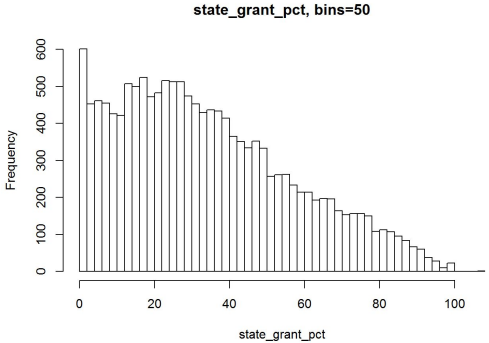
Incomplete Data

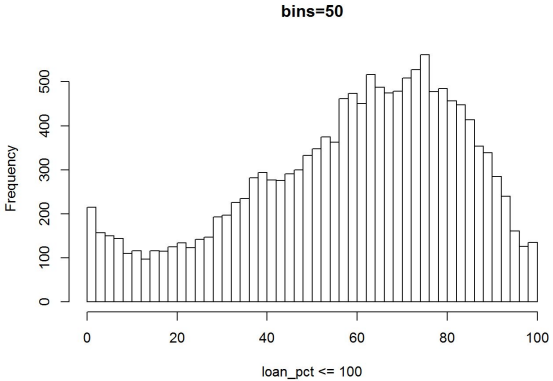
Non-bachelors Institutions

Extreme Outliers, Illegal Values

1417 Colleges and Universities







Aid Model Building

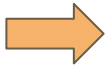
Build up models for both graduation count and graduation rate, trying both “average amount” and “percentage” independent variables

$y_{GradCount} = B_0 + B_1x_{FederalGrantAvgAmt}$

$y_{GradCount} = B_0 + B_1x_{FederalGrantPct}$

$y_{GradRate} = B_0 + B_1x_{FederalGrantAvgAmt}$

$y_{GradRate} = B_0 + B_1x_{FederalGrantPct}$

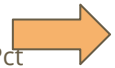


$y_{GradCount} = B_0 + B_1x_{FederalGrantAvgAmt} + B_2x_{StateGrantAvgAmt}$

$y_{GradCount} = B_0 + B_1x_{FederalGrantPct} + B_2x_{StateGrantPct}$

$y_{GradRate} = B_0 + B_1x_{FederalGrantAvgAmt} + B_2x_{StateGrantAvgAmt}$

$y_{GradRate} = B_0 + B_1x_{FederalGrantPct} + B_2x_{StateGrantPct}$



Continue with loan variables

- Decided to federal grants should be the “baseline” to test against since it is most uniform. All schools in USA are affected by same federal grant policy

Limitations to Financial Aid Model

Found that percentage independent variables were best for predicting both graduation count and graduation rate. While statistically significant, results had little practical significance

Besides the small practical effect found, the financial aid model also suffers from an omitted variable issue, just like the cost model

- Also, IPEDS dataset did not have graduation rate broken down by student income level. Did not have same controlling factors as College Scorecard dataset
- Therefore, we we assessed the financial aid model to have an even more severe omitted variable issue and sought a third way

Problem with Both Models - Unobserved Factors

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix}_{n\times k} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_n \end{bmatrix}_{k\times 1} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}_{n\times 1}$$

$$Error\ Term = c + u = u$$ 1

Type of Degree, Quality of Education, Desire of Students to Complete are all included in that error term.

$$E(x_i c_i) <> 0$$ 2

Good schools cost more, so the error term is correlated with our explanatory variables, violating OLS assumptions.

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + c_i + u_{it}$$ 3

So we introduce a time component.

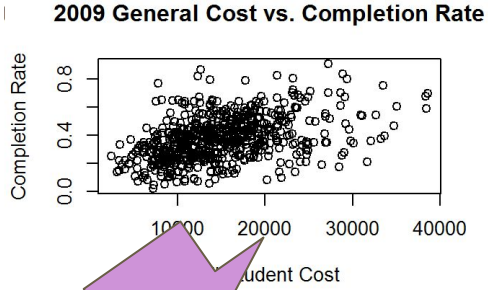
$$y_{it} - y_{i,t-1} = \mathbf{x}'_{it} - \mathbf{x}'_{i,t-1}\ \boldsymbol{\beta} + c_i - c_i + u_{it} - u_{i,t-1}$$ 4

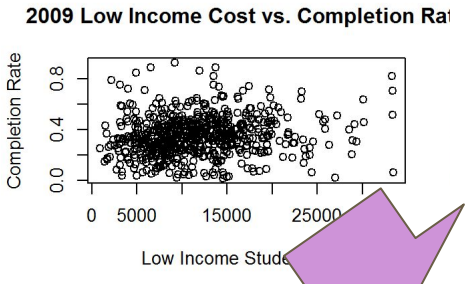
Subtract the difference at T = t and T = t -1

$$\Delta y_{it} = \Delta \mathbf{x}'_{it}\boldsymbol{\beta} + \Delta u_{it}$$ 5

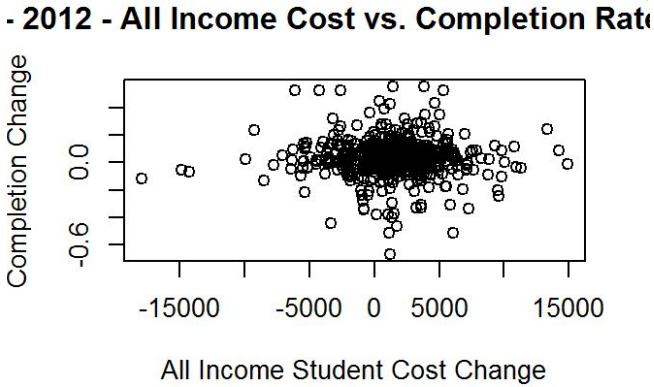
Use OLS to obtain parameter estimates

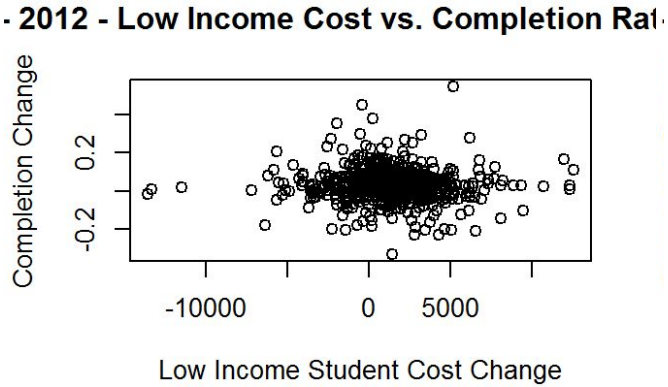
Differences Model





Difference Model is Flat - there no longer appears to be any relationship

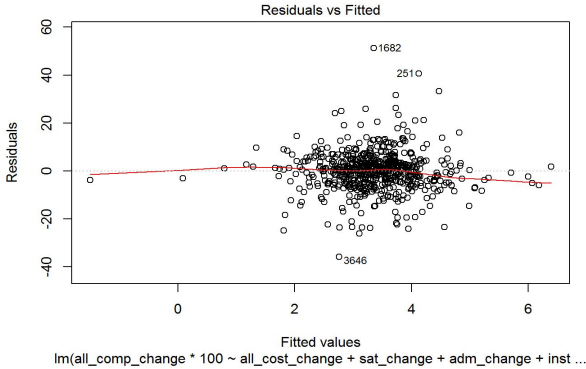




Regression on the Cost Difference Model

```
all_change <- lm(all_comp_change * 100 ~ all_cost_change +
sat_change + adm_change + instrspend_change, data =
analysis_frame)
coeftest(all_change, vcov=vcovHC)
## t test of coefficients:
##
##                       Estimate  Std. Error t value
## (Intercept)         3.29541631  0.39716956  8.297
***
## all_cost_change    -0.00012482  0.00012503 -0.998
## sat_change          0.01499008  0.00813510  1.842
## adm_change         -0.44273447  2.69155615 -0.164
## instrspend_change   0.00021032  0.00019182  1.0964
```

```
low_change <- lm(low_comp_change * 100 ~ low_cost_change +
sat_change + adm_change + instrspend_change, data =
analysis_frame)
coeftest(low_change, vcov=vcovHC)
##
## t test of coefficients:
##
##                       Estimate  Std. Error t value Pr(>|t|)
## (Intercept)         3.5043e+00  6.3722e-01  5.4994 5.55e-08 ***
## low_cost_change    -5.6305e-05  1.7056e-04 -0.3301   0.7414
## sat_change          5.0278e-03  1.2448e-02  0.4039   0.6864
## adm_change         -2.8893e-01  3.6193e+00 -0.0798   0.9364
## instrspend_change   7.9824e-05  2.4674e-04  0.3235   0.7464
```



Conclusions

- Differences model aligns more closely to the OLS key assumptions
- The average cost for all students, and low income students in particular, has no role on institutional completion rates.
- EDA included visual inspections with different categorical control variables and graduation measures:
    - Controls: Percentage of student body in low income category, student population
    - Dependent Variables: Percentage changes in cost
    - Independent Variables: 4 year, 6 year and 8 year graduation time frames
- This result suggests colleges and universities should not attempt incentives or grants as a method to improve completion rates.